TIME COURSE FOR STRUCTURING COMPLEX UTTERANCES

A Thesis
Presented to
The Academic Faculty

by

Christopher M. Crew

In Partial Fulfillment
of the Requirements for the Degree
M.S. in the
School of Psychology

Georgia Institute of Technology
August, 2008

THE TIME COURSE FOR STRUCTURING COMPLEX UTTERANCES

Approved by:

Dr. Zenzi Griffin, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Fredda Blanchard-Fields
School of Psychology
Georgia Institute of Technology

Dr. Dan Spieler
School of Psychology
Georgia Institute of Technology

Date Approved: June 24th, 2008

This thesis is dedicated to the memory of my father Thomas Ray Crew Sr.

ACKNOWLEDGEMENTS

I wish to thank Professor Zenzi Griffin for her patience and guidance and for giving me the opportunity to be a member of her lab. I wish to thank all the faculty, staff and students of the School of Psychology for making my tenure at Georgia Tech a pleasant experience. I wish to thank Professors Fredda Blanchard-Fields and Dan Spieler for their feedback and for sacrificing their time to be on my committee. I wish to thank my mother, my siblings and my nieces and nephews for being my inspiration. Last, and certainly not least, I wish to thank my wife (Nereira) for being my biggest supporter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Eye movements during picture description were used to investigate the time course for structuring embedded clauses. According to the *frame-based* model of language production (e.g., Garrett, 1975; Ferreira, 2000) speakers make decisions about syntactic structure using structural frames corresponding to clauses or verb phrases. *On-line* theories allow structure to be built piecemeal corresponding to individual words and phrases (e.g., Kempen & Hoenkamp, 1987). These predictions were tested in two studies where speakers answered questions based on scenes that depicted someone communicating or thinking about an embedded event, eliciting descriptions like "*A woman is thinking about* (*a man being chased by a bear/a bear chasing a man*)." Based on previous eye movement studies (e.g., Meyer & Van Der Meulen, 2000), gaze shifts between agents (*bear*) and patients (*man*) were expected to occur less often after hearing biased questions, which provided multiple structural cues, than after unbiased ones. The timing of a difference in gaze shifts would then reflect when speakers considered order of mention and committed to an active or passive structure. Study 1 partially supports a frame-based theory of the syntactic planning while study 2 did not provide evidence for either model. Results are discussed in terms of models of language production and reiterate the need for experimental paradigms that use on-line methods in the investigation of the time-course structuring spoken utterances.

CHAPTER 1

INTRODUCTION

> "*Although a speaker may have a general idea of what it is to be said in a sentence before it is produced, the detailed planning for the sentence proceeds in successive phases during each of which a representation of a unit corresponding to a basic clause is constructed and then mapped into a series of words to be uttered* (Ford, 1982, p. 814)."

Such statements have long been at the center of the debate over language production, processing units, and the time course for structuring spoken utterances. One can draw on theories of consciousness as a starting point for providing answers to the questions of when speakers plan their utterance and what units are involved in the planning process. Chafe (1980) described consciousness as the mechanism by which the self makes use of information and is marked by limited capacity, limited duration, and a central and peripheral focus. In many ways this description is analogous to the human capacity to produce and understand language and may prove useful in understanding issues paramount to the study of language production, in particular, the time course for structuring spoken utterances.

Using Chafe's description of consciousness, language production can be said to have limited capacity in the sense that the total amount of information available is vast, but the amount that can be activated at any time is very small in comparison. Second, it has a limited duration in the sense that we are always somewhere in the middle of conscious thought and overt production. Third, language production can be said to have a central focus and a peripheral focus; that is, at any given time a speaker may be actively engaged in some level of the production process (e.g., message formulation) but still have a larger amount of information that to some extent they are aware of (e.g., syntactic structure), but are not “focused” on.

In essence, the study of consciousness reveals a rather mundane yet subtle conclusion about language and our capacity to produce it. The language production process, at its most basic level, can be said to be rule driven, flexible, subject to error, and consisting of units that are, at various points in time, planned and produced. The current study is concerned with the time course for planning syntactic structure but I begin by outlining a model of language production before addressing issues concerning time course.

Language Production

One of the more influential and parsimonious models of language production is that of Levelt (1989). In his seminal book, “Speaking: from intention to articulation”, Levelt provided a detailed account of how speakers go from a communicative intention to articulation. According to Levelt, in conceptualization, speakers conceive an intention and select relevant information from memory or the environment in preparation for the construction of the intended utterance. This is called the message level of representation. Levelt (1989) distinguished between macroplanning and microplanning conceptual processes. Macroplanning involves the elaboration of a communicative goal into a series of subgoals and retrieval of appropriate information. Microplanning involves deciding on matters such as the topic or focus of the utterance. There are two major components of formulation: in one component speakers must select the individual words that they wish to say (lexicalization), and in the other component they have to put them together to form a sentence (syntactic planning). However, theories still differ on the ordering of the two components (e.g., Bock & Levelt 1994; Garrett, 1975).

Implicit in this description of language production is the question of when speakers structure their spoken utterances (e.g., make decisions about syntactic structure)? Given Levelt's architecture, the question of the time course for structuring a spoken utterance is an issue that is addressed in the formulation stage (e.g., lexicalization and syntactic planning). The current study is concerned with the latter. That is, when do speakers make decisions about the syntactic structure of a to-be-produced utterance? To answer this question the current study begins with a description of syntactic structure and theories of how speakers build syntactic structure, moving to a discussion of theories and studies that investigate production levels and units involved in utterance construction, shifting to a discussion of data suggesting that the clause is the major planning unit for syntactic structure, culminating with the introduction of a method that can be used to constrain predictions about the time course for structuring spoken utterances.

CHAPTER 2

LITERATURE REVIEW

Building Syntactic Structure

In many languages, English in particular, the order in which words appear can affect meaning. This ordering of words to get meaning is known as syntax. Syntax is the grammatical structure that determines how words are combined to form sentences (Jay, 2003). For example, "Mickey hit Minnie" does not have the same meaning as "Minnie hit Mickey". This illustrates how impossible it would be to explain the difference in meaning by using a universal rule like, "the actor always comes first and the acted upon comes second", because such a rule would result in misinterpretations of sentences like "Mickey was hit by Minnie." Thus, meaning should be thought of as more than just a string of words because how the string is ordered is also important (Jay, 2003).

Of greater interest to the current study is how this structure is actually created. This question has been addressed by studying the distributional properties of speech errors (e.g., Garrett, 1975). Speech errors are particularly important to the study of language production because they are very systematic and have been found to be useful for speculating about the order of planning processes that precede articulation (e.g., Garrett, 1975; Dell, 1986). One such regularity is that words that participate in exchanges (e.g., the swapping of word order) tend to be from the same syntactic category (Garrett, 1975). For example, when speakers commit a lexical (word) exchange, words from the same syntactic category tend to exchange places. So if a speaker wanted to say, *The dog bit the mailman*, they would be likely to swap the nouns or the noun phrases and come up with, *The mailman bit the dog*, but would be unlikely to say something like, *The mailman dog the bit*. This, along with other categories of speech errors, has been taken to suggest that speakers specify a syntactic plan or frame for a sentence that consist of a series of slots into which content words are inserted and that word exchanges like, *mailman* and *dog*, occur when content words are put in the wrong slot (Garrett, 1975). According to Garrett, that *mailman* and *dog* are able to exchange places, given that they occupy different positions in the syntactic frame, is support for the claim that speakers, to some extent, build syntactic structure prior to speaking. Although Garrett's model has been somewhat controversial it provides a good framework for understanding how speakers might build syntactic structure and has been influential in the development of theories of syntactic processing.

Based on Garrett's findings syntactic processing has been conceptualized as consisting of two stages (*functional* and *positional* processing or simply *grammatical encoding*). Functional processing is where speakers retrieve lemmas (a lexical representation of a word's meaning and its syntactic makeup) and assign grammatical functions such as subject and object (Bock & Levelt, 1994). For example, in an utterance like, *a dog is chasing a mailman*, lemmas for *dog*, *chasing* and *mailman* would be retrieved, and *dog* would be assigned the role of subject and *mailman* would be assigned the role of object. This order of selection commits the speaker to an active sentence structure thus ruling out the possibility of creating a passive sentence structure like *the mailman was chased by the dog* (the selection of active and passive sentence structures will be revisited later). The second component of syntactic processing, positional processing, is concerned with translating grammatical function (e.g., the assignment of the subject—*dog*, and the object—*mailman*) into a linearized component (Ferreira & Engelhardt, 2006). That is, deciding on the order in which words will be placed using a preexisting phrasal frame which contains slots for all the elements of the given phrase (e.g., the determiner *a* and the noun *dog*). Also included within this frame are inflectional affixes (e.g., plurals) so if the word *dog* is plural, the plural morpheme would already be in place and would not have to be separately retrieved and inserted.

In sum, meaning and syntactic structure can be said to be assembled in two stages roughly corresponding to a message and its lemmas followed by assignment of grammatical roles that are fed into a prestored phrasal frame containing all the necessary positional information (Ferreira & Engelhardt, 2006). Missing from this discussion is how or why speakers decide to use a particular syntactic structure (e.g., active or passive) to express their ideas.

There are several possibilities for how this decision might be made. One possibility is that syntactic variation allows the language production system to adapt to differing levels of activation from entities in conceptual memory (e.g., Bock, 1986; Bock & Warren, 1985). This suggests that the most activated concept grabs the earliest syntactic position in the sentence (the subject). Other possibilities are that the earliest sentence position tends to be occupied by nouns that are concrete and imaginable, and animate (e.g., Bock & Warren, 1985; Bock, 1982, respectively). In both instances, something about the message that the speaker wishes to express is influencing syntactic structure and the planning thereof.

The fact that speakers build syntactic structure suggests that speakers engage in some level of planning but in order to address the broader question of the time course for structuring spoken utterances it is important to understand what units are involved in utterance planning. The next section will review theories and research that have addressed the scope of planning and types of production units that may be involved in utterance construction.

Planning Units: Speech Onset Latency, Hesitations, and Pausing Phenomena

Incrementality is assumed to apply to all levels of processing involved in language production (Schriefers, Teruel, & Meinshausen, 1998). This assumption of incremental production has lead researchers to inquire about the size of the increments, or advance planning units, at the different processing levels (e.g., Dell & O'Seaghdha, 1992). A unit can be defined as a basic processing element that can interact and coordinate with other units to form more complex productions. Unit sizes vary with the level of interest. In speech, for example, gestures, phonemes, syllables, words, or clauses could all be considered units based on level of interest. The current study is interested in syntactic planning and will focus on units and theories relevant to syntactic planning (e.g., verbs, clauses, and individual words and phrases).

Scope of Planning: Frames or Individual Words and Phrases?

One explanation for the coordination of words in the production of a sentence is that articulation is initiated when the conceptual planning for a clause (message formulation) but the grammatical encoding (functional and positional processing) for only the first word of that clause has been executed (e.g., Garrett, 1975; Ferreira, 2000). This is known as the *advanced planning* or *frame-based* model of planning in language production. *Frame-based models* are supported by picture-word interference studies that suggest that before speakers start uttering phrases and short sentences containing two names, they select the nouns and the sound form of the first noun (e.g., Meyer, 1996). Meyer presented participants with pictures of pairs of objects that they had to name (e.g., the arrow and the bag) or place in short sentences ("the arrow is next to the bag"). At the same time, the participants heard an auditory distractor that could be related in meaning or sound to the first or second noun, or to both. She found that the time it took participants to initiate speaking was longer when the distractor was semantically related to either the first or second noun, but the phonological distractor only had an effect when it was related to the first noun. This pattern of results suggests that speakers prepare the meaning of short phrases and select the appropriate words before they start speaking, but only retrieve the sound of the first word. An earlier study by Lindsley (1975) also provided support for the frame-based hypothesis. Lindsley asked participants to respond as quickly as possible to a simple picture showing a transitive action (e.g., one person touching another). He found that speakers began phonologically encoding their utterance before they had syntactically encoded the object of the transitive action but not before they knew the verb. Thus, adding further support for the claim that speakers plan at least some portion of a to-be-produce utterance (in this case, up to and including the verb) prior to articulation.

The alternative to advanced planning is an *on-line planning view*. According to this view the planning and production processes can occur simultaneously (e.g., Kempen & Hoenkamp, 1987) and is driven by conceptual information (e.g., Schriefers, Teruel, & Meinshausen 1998). In the Kempen and Hoenkamp model (Incremental Procedural Grammar or IPG) grammatical encoding (functional and positional processing) is completed in such a way that, given a two level process, the first level can release output to the second level to be processed while simultaneously processing another fragment of the utterance in parallel with the processing that level two is doing on the first fragment. This also works for multiple fragments within a single processing stage such that the conceptual planner can release multiple fragments regardless of whether or not the grammatical encoder has finished processing prior conceptual fragments. This does not assume that the difference in the amount processed between conceptual and grammatical encoding is fixed. That is, the conceptual encoder is not always a certain distance ahead of the grammatical encoder. Instead the distance between the two processing stages varies as a function of how quickly processing proceeds at the two stages. This variation is also reflected in the size of the fragment released at each level (e.g., lemma or clause) and the amount of planning completed prior to the onset of speech.

What the Kempen and Hoenkamp model suggests is that the planning process is flexible (incremental). This does not mean that the production system is architecturally incremental, as Wheeldon and Lahiri (1997) suggest, but that speakers may prefer and actually benefit from having variability in syntactic choice. Ferreira (1996) provided support for this model of incrementality. Participants saw the initial frame I GAVE on a computer monitor, and then either the words TOYS followed by CHILDREN or CHILDREN followed by TOYS. In another condition instead of I GAVE participants saw I DONATED. The idea is that if the verb offers flexibility in the order of the arguments following the verb (I gave toys to the children versus I gave children toys) then production will be faster because the system has multiple options. This flexibility is not present with the verb *donate* and thus should take longer to produce if CHILDREN becomes activated sooner than TOYS (e.g., I donated children to the toys is not a grammatical response thus the speaker only has one option). The results supported these predictions. Speakers made fewer errors and had shorter speech onset latencies in the condition that provided syntactic flexibility.

Though informative, these studies are, to some extent, speculative. They can be considered speculative because their conclusions about syntactic complexity and its effect on speech onset latency are based on the observation that speakers are either slower or faster to initiate speech in the presence of a decrease or increase in syntactic complexity or make fewer errors when there is flexibility in syntactic choice. Furthermore, syntactic complexity can be easily confounded with conceptual complexity making it difficult to pinpoint the source of the facilitation (Levelt, 1989). Thus it is arguable that these studies provide an indirect way to determine what speakers are planning by interpreting disfluencies and increases or decreases in speech onset latency, which brings into question any conclusions drawn from such methods. That is not to imply that disfluencies and speech onset latencies are insignificant but that they do not offer a *direct* and *unambiguous* measure of the time course by which speakers make syntactic decisions during articulation, especially in a fluent utterance. Nonetheless, in order to achieve the goal of *directly* addressing the time course for syntactic planning it is important to understand what disfluencies (e.g., pauses and hesitations) have revealed about the planning process before moving to a discussion of how one might *directly* address issues concerning the time course for syntactic planning.

Disfluencies: Hesitations and Pausing Phenomena

According to Goldman-Eisler (1968), disfluencies (e.g., hesitations and pauses) reflect cognitive planning and the complexity of information processing plays a central role in the use and duration of these disfluencies. In a series of studies, Goldman-Eisler (e.g., Goldman-Eisler, 1967, 1968) compared the ratio of silent pauses in speech associated with task of varying levels of difficulty: the description of cartoons versus their interpretation. She found that the cartoon interpretations were correlated with longer pauses than the descriptions, which supports the inference that, for a given task, the longer the delay between stimulus and response, the more cognitive processes are inferred as being required to produce the response. Goldman-Eisler argued that her studies offered evidence that during these pauses speakers planned the content of what they intended to say. Although Goldman-Eisler looked at the relationship between pauses and sentence planning and not syntactic planning (the focus of the current study) her studies provide some evidence that pauses can be used to conjecture about the time-course for syntactic planning, assuming that pauses are necessary for planning to take place. One limitation, as far as this study is concerned, is that pauses might have more than one function. For example, speakers may deliberately use pauses to help the listener segment speech or to give them time to parse speech. Though somewhat problematic, pauses may still be a valuable tool for understanding the time-course for planning syntactic structure.

Other researchers (e.g., Ford & Holmes, 1978; Ford, 1982) looked at the relationship between pauses and clause-level planning. A clause is defined as a grammatical unit that includes, at minimum, a predicate and an explicit or implied subject, and expresses a proposition (Jay, 2003). For example, *it is cold, although the sun is shining,* contains two clauses, *its cold* and *the sun is shining*. Clauses are said to have surface and deep structure. Surface structure refers to the linear ordering that we read or hear while deep structure is the underlying idea or meaning of the surface structure (Medin, Ross, & Markman, 2001). For example, in the sentence, *John preferred to catch the taxi*, the surface structure would be *John preferred to catch the taxi* while the deep structure could be divided into *John preferred it* and *John catch a taxi*. This distinction is important because Ford and Holmes (1978), using a dual-task paradigm, found that reaction time to auditory tones—while simultaneously articulating a sentence—tended to be greater at deep structure clause boundaries which suggest the deep structure clause is an important unit in syntactic planning and that planning can co-occur with articulation.

Another aspect of disfluencies that is of particular significance is their relationship to syntactic complexity. Syntactic complexity affects language processing in many ways. On the comprehension side, an increase in syntactic complexity decreases the acceptability of otherwise grammatical utterances (e.g., Chomsky, 1965), and leads to longer reading times (Gibson, 1998). When it comes to production, the complexity of a message may affect the fluency of the linguistically coded utterance. That is, the initiation time for the production of utterances becomes longer the more complex the syntactic structure of the utterances is (e.g., Ferreira, 1991).

Clark and Wasow (1998) found a relationship between syntactic complexity and disfluencies in spontaneous speech. They found that the more complex a constituent the more likely speakers are to suspend speaking (pause) after an initial commitment to it. They also found that before a hesitant pause, speakers often make a commitment to produce a certain syntactic structure by pronouncing a stranded function word that determines the syntactic category of a phrase; for example, *the* being expressed as *thee*. This suggests that speakers are aware of syntactic complexity and sometimes pause or become disfluent in order to buy time to plan and produce their utterance (see Griffin, 2003 for further evidence that speakers can buffer their speech to adjust for the time need to plan and produce their utterance).

Taken together, these studies predict a relationship between syntactic complexity, disfluencies and syntactic planning. According to Ford (1972), speakers plan syntactic structure basic clause by basic clause, which corresponds to the deep structure (meaning) of the sentence. Furthermore, disfluencies related to syntactic complexity should also be related to planning the deep structure of the clause. This conclusion is based on the finding that processing load increased before basic clause boundaries within sentences and that when speakers pause they did so at the basic clause boundaries.

In sum, although speech errors, speech onset latency and disfluencies have been extremely instrumental in expanding our understanding of the language production process they are characterized by their inability to provide a *direct* measure of the time-course for building syntactic structure. Furthermore, the data from the speech latency studies has been inconsistent and has been reliant on relatively simple utterances. Thus, it is argued that, in order to better understand the time-course for structuring spoken utterances it is necessary to be able gain access to when speakers might plan syntactic structure. Doing so requires the use of techniques (e.g., eye-tracking) and constraints (e.g., syntactic priming) that provide a framework for understanding the time-course for structuring spoken utterances beyond that of speech errors, speech onset latencies, and disfluencies.

The Present Study

Given the review thus far the question becomes, how can one detect when a speaker, especially within a fluent utterance, makes a syntactic decision? To test predictions it is necessary to be able to detect whether a structural decision for a clause may occur after a speaker has begun producing the clause. If speakers may make structural decisions for a clause during speech, it should be possible to detect points during speech when they have not yet decided on which structure to use. Thus, a method is needed that can detect indecision during speech that is related to determining structure but that can co-occur with fluent speech. To this end, the current study will monitor speaker's eye movements during picture description of objects embedded within a scene. It is believed that analyzing speakers eye movements during the description of complex scenes will serve as a better measure of the time course for making syntactic decisions because it lends itself to constraints and predictions that can be directly measured. Furthermore, this method will allow an in-depth look at syntactic planning beyond that of simple noun phrase production.

Determining when speakers make structural decisions would be best accomplished by observing speakers producing sentences of the identical form under conditions where they are either certain or uncertain of which structure to use. Research on syntactic (or structural) priming (e.g., Chang, Bock & Goldberg, 2003), suggests that structural primes facilitate choice of structure when they converge with other constraints on argument order. Syntactic priming refers to the tendency for speakers to repeat a syntactic structure that they have just heard (see Bock 1986, 1989 for first experimental demonstration). Bock (1986, 1989) presented participants with priming sentences that were either prepositional-datives (e.g., *The woman threw a bone to a dog*) or double object-datives (e.g., *The woman threw a dog a bone*), and asked participants to describe pictures like a boy handing a guitar to a musician. The priming was accomplished by having participants repeat either the prepositional-dative or the double object-dative sentence prior to viewing and describing the picture. Bock found that participants primed with the prepositional-dative tended to describe the picture as *The boy handed the guitar to the musician* and participants primed with the double-object-dative tended to describe the picture as *The boy handed the musician the guitar*. This suggests that syntactic priming may arise from the repetition of processes responsible for building syntactic structure during language production (Bock, 1986). Thus, it is argued that syntactic priming may be useful for determining when speakers structure spoken utterances because it provides a mechanism for manipulating the structure of a speakers utterance thereby creating a situation where one can compare speakers behavior (e.g., planning processes) when they are certain and uncertain about the structure of their to-be-produced utterance.

Another potentially useful tool for determining when speakers make structural decisions is the monitoring of eye movements. Several studies have shown a strong correlation between eye movements and word preparation (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998; Meyer & Van Der Meulen, 2000). The general finding is that, when speakers know the locations of objects in a scene they simply move their eyes (shift their gaze) from object to object in the order in which they plan to mention each object (Griffin, 2001). It is certain that speakers shift their gazes between objects in a scene for other reasons besides deciding order of mention. For example, Van Der Meulen (2001) found that speakers make more gaze transitions when there is uncertainty about the number of referents in a scene. This suggest that eye movements are related to other decision making task beside order of mention. Such increases in gaze transitions are though to be related to determining the contents of the message prior to formulation (e.g., Griffin & Bock, 2000). Thus, it may be possible to relate structural indecision to eye movement patterns in the sense that speakers may increase their gaze transitions when they are uncertain about the structure and order of mention of a to-be-produced utterance.

The current study used syntactic priming during picture description to create conditions where speakers were biased to produce a certain syntactic structure and conditions where they were free to make their own decisions about syntactic structure. Biasing the structure of the utterance allows for predictions about when and how structure is selected by comparing speaker's eye movements in instances where the structure is certain (biased) to instances where there is uncertainty about structure (unbiased). Uncertainty refers to not having decided on syntactic structure. This means that speakers have selected the content of their message but have not decided how to put the words together. This was accomplished by having speakers answer questions about pictures of complex scenes (e.g., a woman thinking about a man being chased by a bear/bear chasing a man, see figure 2.1), where *woman* was the experiencer (subject of the main clause), and *man* and *bear* were the patient (acted upon) and agent (actor) of the embedded scene, respectively. The current study uses embedded scenes because they allow for predictions about deciding order of mention and syntax beyond that of single noun phrase production. Furthermore, embedded scenes consist of two clauses that speakers will have to plan, thus creating a situation where predictions can be made about the time course for structuring a more complex utterance.

Figure 2.1 Example of a stimulus for a complex scene depicting a woman thinking about a man (patient) being chased by a bear (agent) or a bear chasing a man (the embedded scene).

It was predicted that speakers would show more indecision about argument order (e.g., who did what to whom) in their eye movements when they produce sentences in an unbiased condition rather than a biased one. This prediction is based on previous eye movement studies showing that speakers make more gaze transitions when there is uncertainty about the number of referents in a scene (Van Der Meulen, 2001). Using uncertainty as an index of decision-making makes it possible to formulate predictions about eye movement patterns expected for both the on-line and frame-based theories of production. According to the one-line view, speakers may start their sentences without having decided the order of mention and subsequently syntactic structure. In contrast, the frame-based view predicts that structural decisions could take place earlier. As a result, speakers may only show signs of being unsure of which structure to use prior to speech. This suggests that eye movements (hereafter referred to as gaze shifts) will be most frequent prior to speech onset. Stated differently, if speakers make decisions about the syntactic structure of an embedded clause prior to speaking they will make a greater number of gaze shifts between the agent (*bear*) and patient (*man*) (or vice versa) prior to speaking in the unbiased condition. Alternatively, if speakers plan the structure of an embedded clause piecemeal they still must select the information for the first clause (e.g., lemmas for nouns and verbs) before they start uttering the second clause, but they can make decisions about the structure of the embedded clause during articulation of the main clause. This would be evidenced by an increase in the number of gaze shifts between the agent (*bear*) and patient (*man*) (or vice versa) in the unbiased condition during articulation of the main clause with the greatest number of shifts occurring as late as the onset of the main verb.

The current study focused on five points in time as indices of the time course for planning an embedded clause. Of primary interest is the frequency of gaze shifts between the embedded subject (*bear*) and the embedded object (*man*) (or vice versa) during the points in time corresponding to: 1) the onset of speech, 2) the onset of the subject of the main clause (*woman*), 3) the onset of the main verb (*thinking*) 4) the onset of the embedded clause, and 5) the onset of the embedded subject (*bear*) or embedded object (*man*) of the embedded clause (see figure 2.2). These five points in time were chosen because they correspond to the content words of the sentence. Content words are important because they carry the main information load and meaning in a sentence thus representing critical components of the sentence that would have to be planned (e.g., order of mention and syntactic structure).

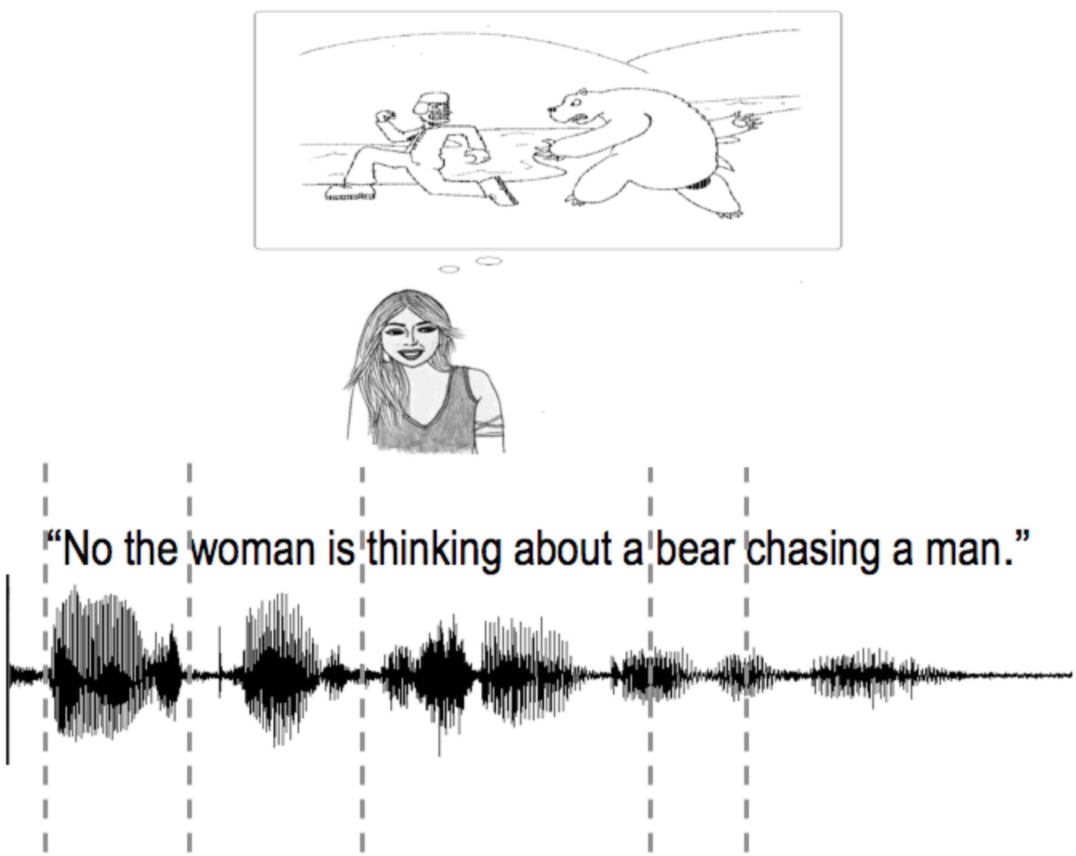

Figure: 2.2. The points in time of interest for the response, "No a woman is thinking about a bear chasing a man (or a man being chased by a bear)", are as follows: 1) the onset of speech (*no*), 2) the onset of the subject of the main clause (*woman*), 3) the onset of the main verb (*thinking*), 4) the onset of the embedded subject (*bear*) or object (*man*) noun phrase (or vice versa), and 5) the onset of the embedded verb (*chasing*).

CHAPTER 3

STUDY 1

Method

Participants

Twenty-four undergraduates from the Georgia Institute of Technology participated in the experiment for course credit. The data of twenty-three participants are reported here. The data of one participant was excluded because of an overwhelming tendency to use more active than passive sentence structures (23 active versus 1 passive). This bias suggests that this particular participant had not been affected by the priming manipulation. All participants were native speakers of American English, 18-23 years old, and had normal or corrected-to-normal vision.

Apparatus

Eye movements were monitored using an ISCAN (model number ETL-400) tabletop eye-tracker. Stimuli were displayed on a 21-in. monitor. Speech was digitally recorded at 12 kHz using a headset microphone. Participants placed their foreheads against a rest to minimize variation in participant distance from the computer monitor (approximately 61-cm).

Materials

Stimuli consisted of 48 line drawings of multiple objects (e.g., people, animals, trucks) involved in some sort of action, embedded within a scene. Twenty-four of the drawings were the critical scenes and the other 24 were fillers. The critical scenes depicted a person thinking about, laughing at or reading about events, each with a visible experiencer, agent, and patient, such as a woman (experiencer) looking at a bear (agent) chasing a man (patient) (see figure 2.1). The fillers showed a person thinking about, laughing at, or reading about a mixture of transitive and intransitive events, such as a boy reading about a woman playing a guitar. A set of 96 questions was also recorded (twenty-four active questions, twenty-four passive questions, twenty-four neutral questions, and 24 fillers with an active sentence structure). Seventy-two of the questions were used for the critical scenes and 24 were used for the filler scenes. There were three questions recorded for each picture, two biased (active, passive) and one unbiased (neutral). The questions for the critical scenes either biased an active sentence structure ("*Is a woman thinking about something chasing someone?*"), a passive sentence structure ("*Is a woman thinking about someone being chased by something?*"), or was unbiased ("*Is a woman thinking about something happening?*") (see Table 3.1). The questions for the fillers scenes were similar to the questions for the critical scenes but in the filler scenes the question provided all the grammatical and structural information (for example the nouns, verbs, and structure are provided in the filler question, "*Is a man laughing at donkey playing basketball?*").

Table 3.1. Examples of biased (active, passive) and unbiased (neutral) questions

Condition	Prime	Examples
Biased	Active	Is a woman thinking about something chasing someone?
Biased	Passive	Is a woman thinking about someone being chased by something?
Unbiased	Neutral	Is a woman thinking about something happening?

Design

This study used a one-factor, within-subject design with the independent variable being structure of the question (biased vs. unbiased) and the dependent variable being the gaze shifts between embedded subject (*bear*) and the embedded object (*man*) (or vice versa) in the embedded scene. The critical and filler pictures were the same across all trials. The only thing that varied was the question (biased or unbiased). The questions were counter balanced across 3 lists using a Latin Squares design, such that each participant was exposed to each level of the independent variable. Each session consisted of 48 trials (hearing a question, seeing a picture and responding to it). The fillers were randomly dispersed throughout the lists. None of the items were reused for experimental and filler displays.

Procedure

Participants answered questions about pictures containing cartoon drawings of events. Before each picture appeared on the screen they heard a question about it (the interstimulus interval between hearing the question and seeing the picture was 500 milliseconds). Participants were instructed to answer the question by giving a full description of what they saw happening in the scene using a complete sentence. It was conceivable that participants could answer the questions using a one-word answer such as *"yes"* or *"no"* so the instructions were written such that participants were aware that one-word answers were considered incomplete. It was critical that participants used complete sentences because the current study is concerned with the time-course for structuring spoken utterances and a one-word answer would have restricted our ability to test predictions. Thus, if the answer was *"yes"* participants were permitted to answer *"yes"* but they had to then give a full description of the scene. Conversely, if the answer was *"no"* participants could have answered *"no"* but they also had to give a full description of the scene. The critical trials were always *"yes"* answers and the filler trials were always "*no*" answers.

Coding

Voice recordings were transcribed. Trials in which participants did not follow the prime (12.9%), did not mention the experiencer, agent or patient (10.1%), or did not produce reversible utterances (11.1%), were excluded. Reversible utterances were utterances that had a "by" separating the embedded subject from the embedded object. For example, a response like, *a woman is thinking about a man being chased by a bear* is reversible (reverse to *bear chasing a man*) where a response like, *a woman is thinking about a man running from a bear* is not. Reversing the argument order of the latter example would change the meaning (*man running from a bear* and *bear running from a man*). Of the 768 trials, 586 met criteria for inclusion. Sphinx-2 (CMU Sphinx Group, 2001) was used to force align speech recordings with transcriptions of the speech recordings. Sphinx-2 supplied measurements (in milliseconds) of onsets of speech, subject and object nouns, and the main and embedded verbs for each picture description. These measurements were later used to match the eye data with the speech data.

All eye data was recorded as screen coordinates (e.g., using 800 x 600 pixels). Saccades were defined as sequences of sampled fixation coordinates spanning at least 0.5° of visual angle, for which the velocity of movement exceeded 40° per second, beginning and ending when velocities fell below 20° per second. Remaining samples constituted fixations if they had a minimum duration of 50 milliseconds. For each scene, the computer program Matlab was used to draw polygons around the experiencer, agent and patient with a margin of approximately 1° of visual angle to identify fixated regions. Gazes were defined as beginning at the onset of a fixation within a polygon and ending with offset of the last fixation within it. Thus, the duration of saccades within a polygon contributed to gaze durations (see Irwin, 2004). Once the polygons were defined for each picture, the data files were processed to create fixations and saccades. After the fixations and saccades were determined the fixation files were run through another routine that labeled the locations of the fixations and collapsed the fixations into gazes (which began when the eye entered a polygon and ended when it left the region regardless of the number of fixations within the polygon). Time spent gazing at the experiencer, agent, and patient prior to the onsets of speech, the subject and object nouns, and verbs was calculated for each trial and region. From these values, mean gaze times per time period and condition were calculated for participants and items. The mean number of gaze shifts between the agent and the patient (or vice versa) was the main dependent measure.

Results

The analyses consist of t-test using matched pairs that compare the number of gaze shifts between the agent and patient, the duration of speech, speech onset latency, total gaze duration and the time spent looking at a particular referent before mentioning it in both the unbiased and biased conditions. Furthermore, the data have been analyzed across subjects and across items. The analyses across items either did not reach statistical significance or failed to reveal a different pattern of results than the by-subject analyses so they are not presented in this discussion.

Gaze Shifts

All results are represented graphically in *figure: 3.1.*

The analysis of the gaze shifts for each point in time revealed only one statistically significant comparison. There was no difference in the number of gaze shifts between the agent and patient (or vice versa) in the biased and unbiased conditions for the points in time corresponding to 1) the onset of speech, 2) the onset of main verb (*thinking*), 3) the onset of either the embedded subject (*bear*) or object (*man*) of the embedded clause, or 4) the onset of the embedded verb (*chasing*). There was a statistically significant difference in the amount of gaze shifts between the agent and patient (or vice versa) for the point in time corresponding to the onset of the main subject (*woman*). During this point in time the amount of gaze shifts between the embedded subject and object was greatest for the unbiased (M = .27) in comparison to the biased (M = .17) condition $t(31) = 2.39$, $p < .05$ (see table 3.2).

In order to make better sense of these results, in terms of their relevance to the time course for structuring a spoken utterance, it is important to analyze other variables (e.g., speech onset latency, gaze duration and disfluencies) that may have an impact on the pattern of results.

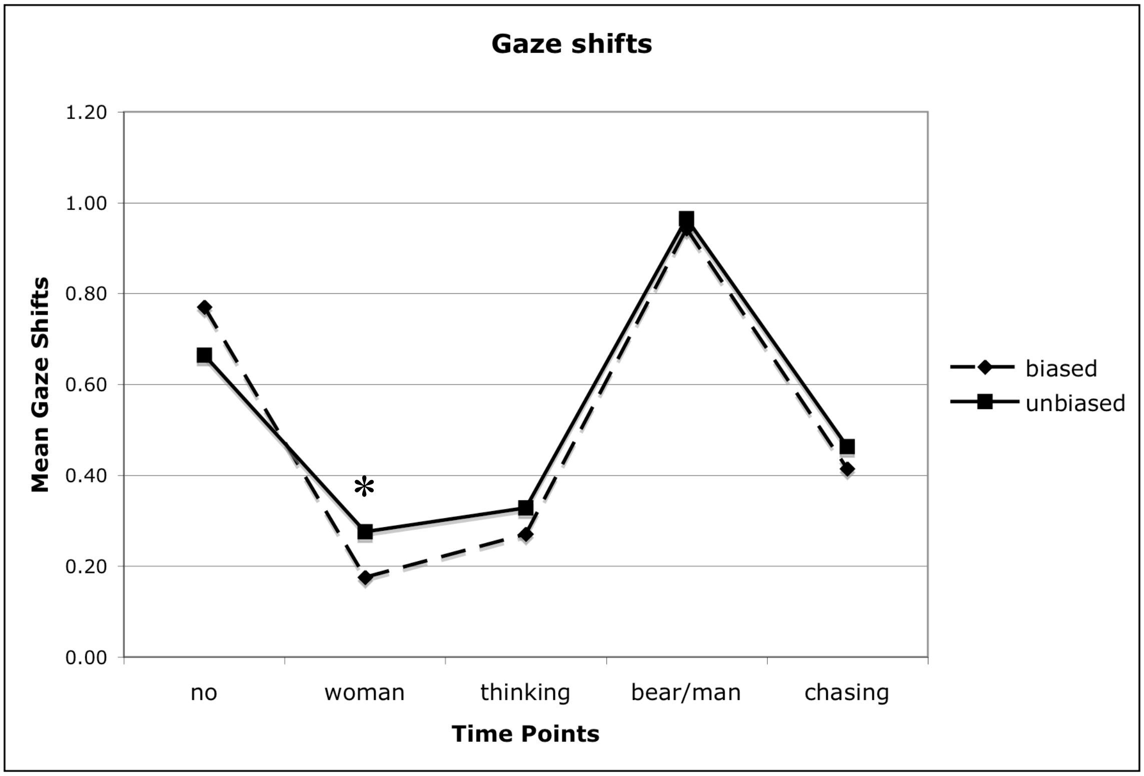

Figure: 3.1. Graph of the mean number of gaze shifts between the agent and patient (or vice versa) for biased and unbiased trials during the points in time corresponding to the onset of 1) speech, 2) the main subject, 3) the main verb, 4) the embedded subject or object of the embedded clause (or vice versa), and 5) the embedded verb. (Note: * = significant at $p < .05$).

Table 3.2. Mean (and standard error) of gaze shifts per time point as a function of bias

Time Points	Biased	Unbiased	T-Value	P-Value
Speech Onset	.77(.08)	.66(.08)	-1.28	p > .05
Main Subject	.17(.04)	.27(.04)	2.39	p < .05
Main Verb	.27(.05)	.32(.05)	1.34	p > .05
Embedded Subject/Object	.94(.05)	.96(.05)	.38	p > .05
Embedded Verb	.41(.05)	.46(.05)	.87	P > .05

Disfluencies

All results are represented graphically in *figure: 3.2.*

The analysis of disfluencies in speech looked at the difference in the occurrence of disfluencies as a function of bias. Disfluencies are of particular interest because they suggest problems in the timing of word retrieval and might assist in understanding the gaze shift results. For this analysis, two independent raters used a set of rules to code the transcribed speech for possible disfluencies. A disfluency was defined as a noticeable pause over 200 ms during speech that was not a natural pause (e.g., occurring at a phrase or clause boundary) or was longer than expected at a boundary, stressed articles, so-called fillers or filled pauses such as *UM* and *UH*, a prolonged or stretched out word, a repeated word, or a cutoff word. The percent agreement and the inter-rater reliability for the raters for each of the five time periods of interest consistently fell within the good to excellent range (Cohen's Kappa of .572 and .863, respectively).

The disfluency analysis looked for differences in disfluencies before content words. The content words corresponded to 1) the period of time prior to the onset of first content word (no), 2) between the first content word (no) and the main noun (*woman*), 3) between the main noun (*woman*) and the main verb (*thinking*), 4) between the main verb (*thinking*) and the subject noun phrase (*bear*) or object noun phrase (*man*) of the embedded clause (or vice versa) and, 5) between the subject noun phrase (*bear*) or object noun phrase (*man*) of the embedded clause and the embedded verb (*chasing*). Of the five periods of time, three reached statistical significance but only one was in the expected direction (e.g., unbiased greater than the biased condition). The mean number of disfluencies for the period of time between the first content word and the main noun (*woman*) was greater in the biased (M = .160) than the unbiased (M = .097) condition *t*(31) = -2.07, p < .05. The mean number of disfluencies for the period of time between the main noun (*woman*) and the main verb (*thinking*) was greater in the biased (M = .059) than the unbiased (M = .023) condition *t*(31) = -2.32, p < .05. Finally, the mean number of disfluencies for the period of time between the main verb (*thinking*) and the subject noun phrase of the embedded clause (*bear*) was greater in the unbiased (M = .174) than the biased (M = .133) condition *t*(31) = 2.12, p < .05 (see table 3.3).

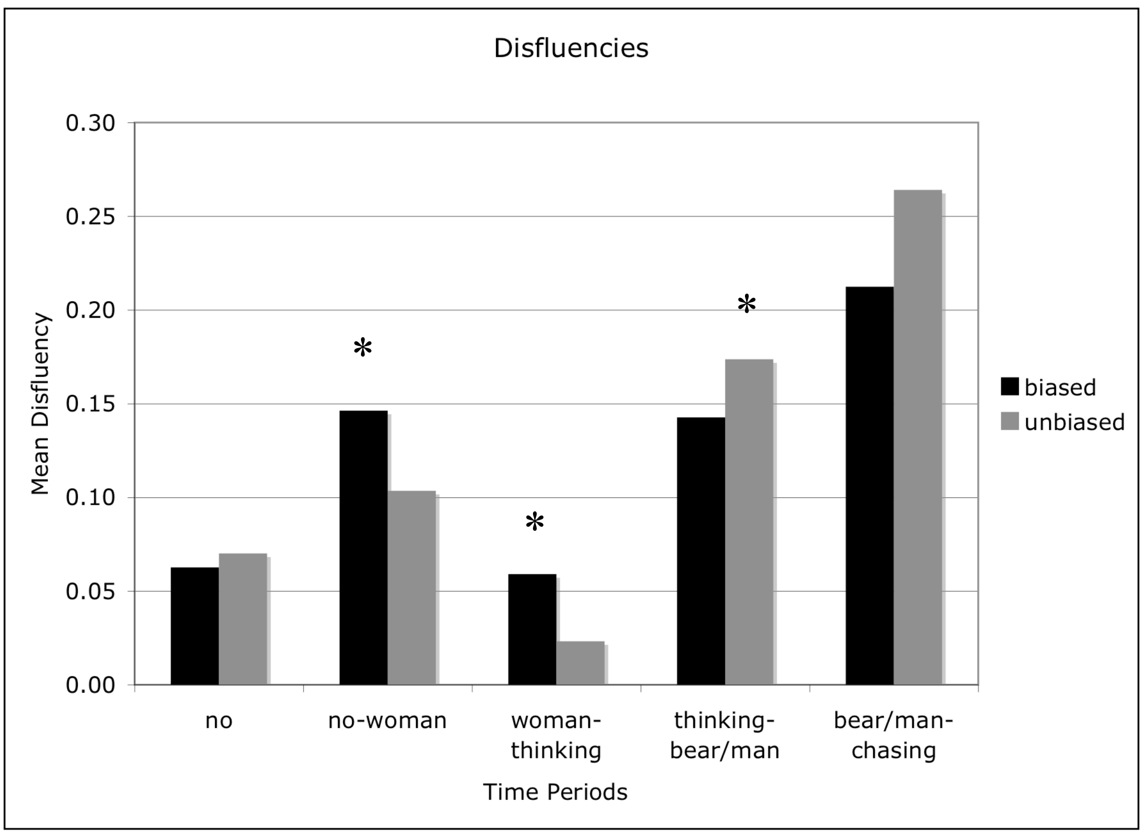

Figure: 3.2. Graph of the occurrence of disfluencies for biased and unbiased trials during the periods of time corresponding to: 1) prior to the first content word 2) between the first content word and the main noun, 3) between the main noun and the main verb 4) between the main verb and the embedded subject or object, and 5) between the embedded subject or object and the embedded verb. (Note: * = significant at $p < .05$).

Table 3.3. Mean disfluencies (and standard error) per time period as a function of bias

Time Periods	Biased	Unbiased
Before first content word	.07(.01)	.07(.01)
First content word to main subject	.16(.03)	.09(.03)
Main subject to main verb	.05(.01)	.02(.01)
Main verb to embedded subject/object	.13(.01)	.17(.01)
Embedded subject/object to embedded verb	.20(.03)	.25(.03)

Speech Duration

The analysis of speech duration tested for a difference between the speech duration corresponding to periods of time before specific words (e.g., content words) to see if there was an effect of bias on the amount of time it took speakers to articulate their utterance in reference to content words. This analysis was concerned with the speech duration between four periods of time 1) between speech onset and the onset of the main noun (*woman*), 2) between the onset of the main noun and the onset of the main verb (*thinking*), 3) between the onset of the main verb and the onset of the head noun of the embedded subject noun phrase (*bear*) and, 4) between the onset of the main verb and the onset of the head noun of the embedded object noun phrase (*man*). This was accomplished by subtracting the duration of the current period of time of interest from the previous period of time of interest. For example, speech duration for period of time between the onset of the main verb (*thinking*) was calculated by subtracting the total speech duration up to the onset of the main verb from the total speech duration up to the onset of the main noun (*woman*). This analysis revealed no statistically significant results.

Speech Onset Latency

The analysis of speech onset latency looked at the difference in speech onset as function of bias. This was done to determine if speakers took longer to initiate speech when they were not primed with a specific sentence structure. The idea is that, when forced to come up with the content of the to-be-produced utterance speakers should delay their speech as they gather the necessary information (e.g., lemmas and argument order). This analysis revealed no statistically significant results.

Gaze Onset and Gaze Duration

The analyses for gaze onset and gaze duration looked at the difference in the initial gaze onset and the gaze duration before specific referents as a function of bias. The goal was to see if there was a difference in the initial gaze onset or the amount of time spent looking at a particular referent (gaze duration) before mentioning it (also suggests retrieval difficulty) as a function of bias. Gaze duration was analyzed for the points in time corresponding to 1) immediately prior to mentioning the main subject (*woman*), and 2) immediately prior to mentioning the second noun (*man* or *bear*, depending on whether the sentence was active or passive). These analyses revealed no statistically significant results.

Discussion

The mean number of gaze shifts between the agent and patient was greater for unbiased condition in comparison to the biased condition only for the point in time corresponding to the onset of the main noun (*woman*). This pattern of results is consistent with the frame-based model of syntactic planning (e.g., Garrett, 1975, Ferreira, 2000) where speakers structure their utterances either prior to articulation or very early during articulation (before mentioning the verb). This pattern of results is also consistent with the predictions about indecision and eye movements in the sense that indecision or uncertainty about the structure of the to-be-produced utterance was evident only during the early stages of utterance construction (e.g., during articulation of the main subject, *woman*). That is, it is arguable that speakers showed evidence of not having decided on a structure and thus made more gaze shifts between the agent and patient (or vice versa) in the unbiased condition where there was less structural information available.

Although there is some evidence that speakers were able to put off making decisions about the embedded structure (at least until they mentioned the main subject) it is unclear whether speakers were actually making structural decisions or simply determining the truth-value of the question. That is, all the critical trials required a *yes* response and all the filler trials required a *no* response so it is unclear whether the pattern of gaze shifts is a result of speakers actually planning the structure of their utterance or speakers making gaze shifts in order to confirm that the answer to the question was either *yes* or *no*. This question is raised in light of the considerable amount of gaze shifts prior to speech onset (suggest that speakers where engaged in some level of processing (e.g., message formulation or determining the truth-value) prior to speaking). Thus, in an attempt to increase the likelihood that speakers eye movements were actually reflecting decisions about the structure of the to-be-produced utterance study 2 was designed such that the answer to all questions was always no, hopefully increasing the likelihood that the gaze shifts reflected decisions about syntactic structure and effectively reduce the number of gaze shifts prior to speech onset. Furthermore information in the filler question and the picture (e.g., subject, verb and object) and the main subject, main verb and embedded verb of the critical question and picture matched only a portion of the time. This was done in order to decrease the monotony of the task and to encourage variability in responses.

Another change was in the location of the subject and object in the filler and critical pictures. In study 1 the agent and patient (in the filler and critical trials) were always in the same location (either left or right). In study 2 the location of the agent and patient (in the filler and the critical trials) was varied. This was done to increase the likelihood that speakers actually had to move their eyes to gather information about the subject and object of the embedded scene and subsequently make decisions about order of mention and syntactic structure. Finally, the names of the main subject (*woman*) and the main verb (*thinking*) were varied. In study 1 the main subject was disproportionately described as *a man* or *a woman* and the main verb phrase was disproportionately described as *thinking about*, *looking at* or *talking about*. The main subject was changed to more concrete labels of the animate nouns (e.g., *doctor or lawyer*) and the verb was changed in order to increase variation in responses (e.g., *laughing*, *singing*, *crying, etc.)*.

CHAPTER 4

STUDY 2

Method

Participants

Thirty-two undergraduates from the Georgia Institute of Technology participated in the experiment for course credit. All participants were native speakers of American English, 18-23 years old, and had normal or corrected-to-normal vision.

Apparatus

The same equipment from study 1 was used to record eye movements and speech in study 2.

Materials

The same pictorial stimuli from study 1 were used for study 2, with the exception of 4 pictures that were replaced because participants in study 1 seemed to have trouble describing the embedded scene (e.g., a boxer punching a boy was described as a boxing match). There was also a slight change in the structure of the questions. This change was made in response to the assumption that the results in study 1 may be due to the experimental design reflecting the need for speakers to determine the truth-value of the question before they can respond. Thus changes were made to correct for this possibility.

A set of 96 questions was also recorded (twenty-four active questions, twenty-four passive questions, and twenty-four neutral questions). Seventy-two of the questions were used for the critical scenes and 24 were used for the filler scenes. There were three questions recorded for each picture, two biased (active, passive) and one unbiased (neutral). The questions for the critical scenes either biased an active sentence structure (*Is a lawyer thinking about something chasing someone?*), a passive sentence structure (*Is a lawyer thinking about someone being chased by something?*), or was unbiased (*Is a lawyer thinking about something happening?*) (see table 3.1). The questions for the filler scenes were similar to the questions for the critical scenes but in the filler scenes the question provided all the necessary information (e.g., the subject, verb, and object were provided in the question, "*Is a man laughing at donkey playing basketball?*).

Design

The design for study 2 was identical to the design used in study 1.

Procedure

Just as in study 1, participants answered questions about pictures containing cartoon drawings of events. Before each picture appeared on the screen they heard a question about it (the interstimulus interval between hearing the question and seeing the picture was 500 milliseconds). Participants were instructed to answer the question by giving a full description of what they saw happening in the scene using a complete sentence (see study 1 for further explanation of a "complete" answer).

Coding

The coding strategies and criterion for the eye and speech data were identical to the coding strategies and criterion used in study 1.

As in study 1, voice recordings were transcribed. Trials in which participants did not follow the prime (14.2%), did not mention the experiencer, agent or patient (1%), or did not produce reversible utterances (11.2%) were excluded. Of the 768 trials, 587 met criteria for inclusion.

Results

The analyses consist of t-test using matched pairs that compare the number of gaze shifts between the agent and patient, the duration of speech, speech onset latency, total gaze duration and the time spent looking at a particular referent before mentioning it in both the unbiased and biased conditions. The analyses for speech onset, gaze onset and gaze duration did not produce any statistically significant results and are not presented here. Similar to study 1, the analyses across items either did not reach statistical significance or failed to reveal a different pattern of results and are not presented in this discussion.

Gaze Shifts

All results are represented graphically in *figure: 4.1.*

The analysis of the gaze shifts for each point in time revealed no statistically significant results, (see table 4.1.). Overall, the amount of gaze shifts between the agent and patient was greatest during the points in time corresponding to the onset of the embedded clause (the relevance of this finding is discussed in greater detail in the discussion section).

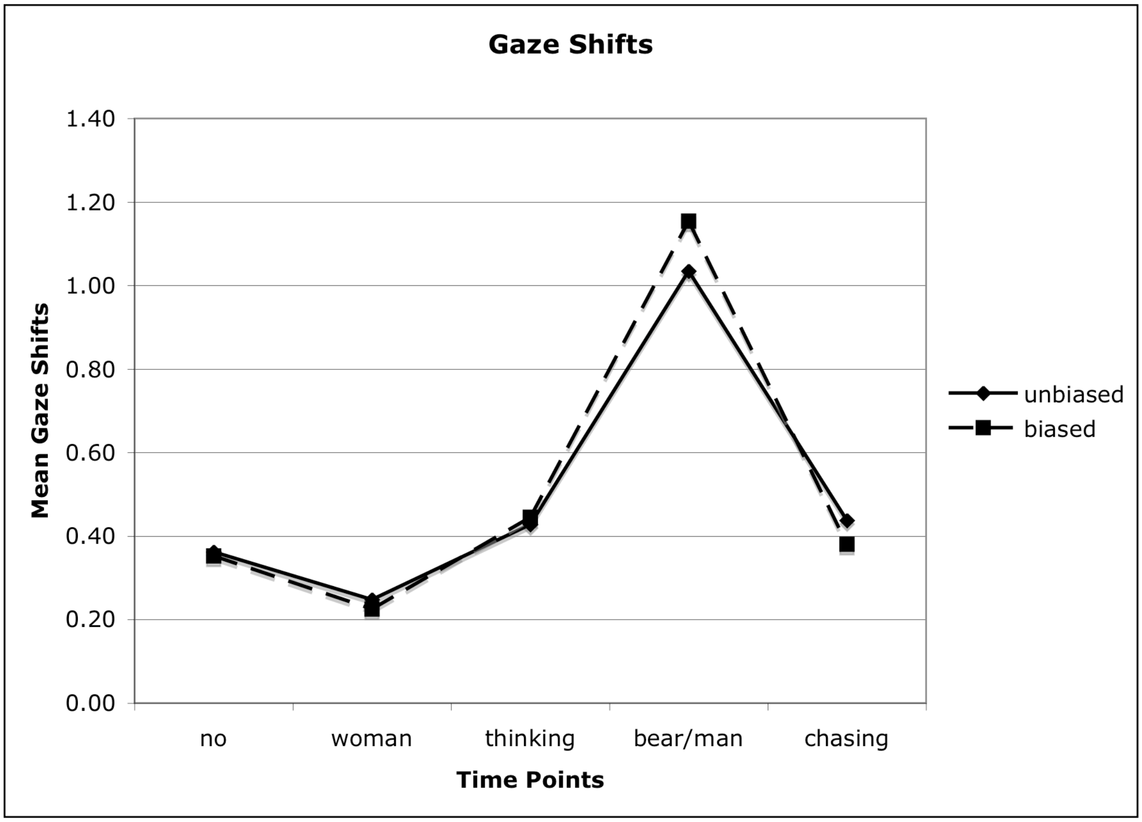

Figure: 4.1. Graph of the mean number of gaze shifts between the agent and patient for biased and unbiased trials during the time points corresponding to the onset of 1) speech, 2) the main subject, 3) the main verb, 4) the embedded subject or object of the embedded clause (or vice versa), and 5) the embedded verb.

Table 4.1. Mean (and standard error) of gaze shifts per time point as a function of bias

Time Points	Biased	Unbiased	T-Value	P-Value
Speech Onset	.35(.04)	.36(.04)	.22	p > .05
Main Subject	.22(.03)	.24(.03)	.61	p > .05
Main Verb	.44(.05)	.42(.05)	-.34	P > .05
Embedded Subject/Object	1.15(.08)	1.03(.08)	-1.39	p > .05
Embedded Verb	.38(.05)	.43(.05)	1.06	P > .05

Disfluencies

All results are represented graphically in *figure: 4.2.*

As in study 1, two independent raters used a set of rules to code the transcribed speech for possible disfluencies. The percent agreement and the inter-rater reliability for the raters for each of the five time periods of interest (same as study 1) consistently fell within the good to excellent range (Cohen's Kappa of .572 and .863, respectively).

Similar to study 1 the disfluency analysis looked for differences in disfluencies before content words corresponding to the periods of time 1) prior to speech onset 2) between speech onset and the onset of the main noun (*woman*), 3) between the onset of the main noun (*woman*) and the onset of the main verb (*thinking*), 4) between onset of the main verb (*thinking*) and the onset of the embedded subject noun phrase (*bear*) or the embedded object noun phrase (*man*) of the embedded clause and, 5) between the onset of the embedded subject noun phrase or the embedded object noun phrase of the embedded clause and the onset of the embedded verb (*chasing*). Of the four periods of time, one reached statistical significance but was not in the expected direction (the biased condition was greater than the unbiased condition). That is, the mean number of disfluencies for the period of time immediately prior to the onset of the first content word (*no*) was greater in the biased (M = .144) than the unbiased (M = .058) condition $t(31) = -2.80$, $p < .05$ (see table 4.2).

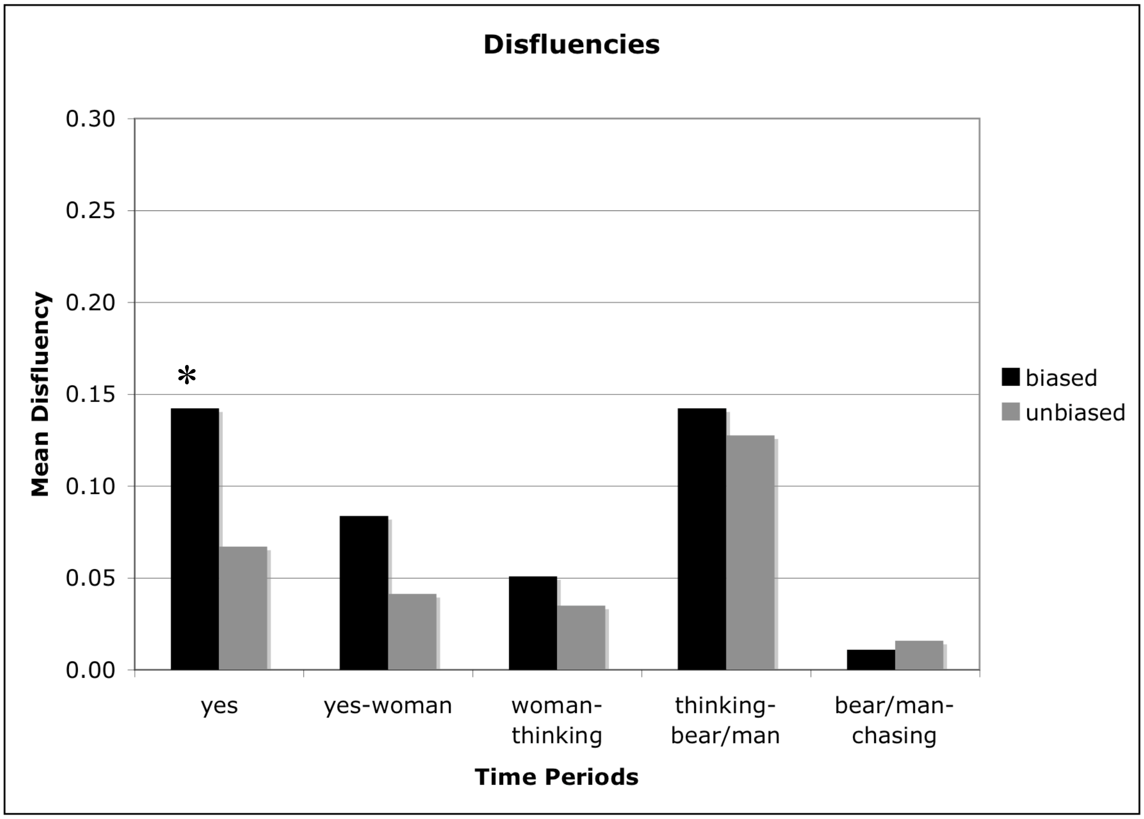

Figure: 4.2. Graph of the occurrence of disfluencies for biased and unbiased trials during the periods of time corresponding to 1) immediately prior to the first content word 2) between the onset of the first content word and the onset of the main subject, 3) between the onset of the main subject and the onset of the main verb 4) between the onset of the main verb and the onset of the embedded subject or object, and 5) between the onset of the embedded subject or object and the embedded verb. (Note: * = significant at $p < .05$).

Table 4.2. Mean disfluencies (and standard error) per time period as a function of bias

Time Periods	Biased	Unbiased
Before first content word	.14(.02)	.06(.02)
First content word to main subject	.08(.02)	.04(.02)
Main subject to main verb	.04(.01)	.03(.01)
Main verb to embedded subject/object	.14(.02)	.12(.02)
Embedded subject/object to embedded verb	.01(.01)	.01(.01)

Speech Duration

As in study 1, this analysis was concerned with the speech duration prior to four periods of time 1) between speech onset and the onset of the main noun (*woman*), 2) between the onset of the main noun and the onset of the main verb (*thinking*), 3) between the onset of the main verb and the onset of the head noun of the embedded subject noun phrase (*bear or man depending on the sentence structure*) and, 4) between the onset of the main verb and the onset of the head noun of the embedded object noun phrase (*bear* or *man depending on the sentence structure*) (see study 1 for discussion of calculations). The overall duration of speech (time taken to articulate their utterance) was greater in the biased (M = 8308.02 ms) than in the unbiased (M = 7954.74 ms) condition $t(31) = -1.96$, $p < .05$. The speech duration for the period of time between the onset of speech and the onset of the main noun (*woman*) was greater for the biased (M = 832.196 ms) than the unbiased (749.044 ms) condition $t(31) = -2.25$, $p < .05$. The speech duration for the period of time between the onset of the main verb (*thinking*) and the onset of the head noun of the embedded subject or object noun phrase (depending on the structure) was also greater for the biased (M = 1982.31 ms) than the unbiased (M = 1833.89 ms) condition $t(31) = -2.82$, p < .05).

Discussion

The pattern of gaze shifts between the agent and patient (or vice versa) for the biased and unbiased conditions failed to support the predictions or replicate finds of previous eye movement studies showing that gaze shifts between referents are greatest when there is uncertainty in the structure of a to-be-produce utterance (e.g., Van Der Meulen, 2001). This lack of an effect of bias suggests that the manipulation was ineffective (e.g., participants did not consistently make use of the structural information provided in the primes) and that the gaze data simply reflects time needed to prepare a single-word utterance. In other words, the gradual increase and sudden spike in gaze shifts just prior to the onset the embedded clause may reflect the fact that speakers tend to look at objects just before they mention them. This possibility is supported by previous eye movement studies (e.g., Meyer, Sleiderink, & Levelt, 1998; Griffin & Bock, 2000) that show that speakers look at objects they plan to mention about 1000 ms before they mention them. According to Griffin (2001) and Meyer et al. (1998) the 1000 ms reflects the time necessary to select a word and retrieve its pronunciation. Given that this is the pattern seen in the gaze shift data in study 2 (a sudden spike in gaze shifts about 1000 ms prior to mentioning the agent or patient), that the gaze shifts were time locked to the onset of the agent or patient (depending on the sentence structure), and there was no difference in the number of gaze shifts between the agent and patient (or vice versa) at any point in time, it is unclear whether study 2 can be taken as support for clause level planning or indecision about order of mention and syntactic structure.

CHAPTER 5

GENERAL DISCUSSION

The primary issue addressed in this study is the extent to which speakers make decisions about syntactic structure prior to or during articulation (it is important to note that, for illustrative purposes, the representation of both views has focused on the more extreme planning possibilities. But many hold the view that the planning process encompasses both on-line and frame-based components). One hypothesis, *on-line planning*, states that speakers can make decisions about order of mention and syntax piecemeal, that is, during articulation (e.g., Kempen & Hoenkamp, 1987). This view allows for flexibility in the selection of syntactic structure by allowing speakers to structure their utterance in parallel with articulation. Another hypothesis, *frame-based planning*, states that speakers make decisions about syntax and order of mention prior to speaking (e.g., Garrett, 1975; Ferreira, 2000). This approach reduces the flexibility of the planning and production process because speakers have to plan a fixed amount of information before they can begin articulation.

Given the equivocal pattern of results, the best answer to the question of the time-course for structuring spoken utterances is probably that it depends. It is highly unlikely that there is a set of situational constraints that call for *advanced* or *on-line planning* that can be generalized across speakers, situations, and utterances. However, it is likely that speakers have preferences and possibly even schemas for minimizing or maximizing preparation under various circumstances (e.g., a high working memory demand). The latter possibility has been explored in previous online studies of language production (e.g., Ferreira & Swets, 2001). Ferreira and Swets (2001) used speech onset latency during arithmetic calculation (e.g., the sum of 14 and 7 is X) to show that the extent to which planning occurs is at least partially under the control of the speaker, and depends on the intentions that motivate speech.

Summary of Results

The prediction that the difference in gaze shifts between the agent and patient (or vice versa) would be greatest in the unbiased condition and would correspond to the selection of words and individual phrases (on-line incremental planning) was not supported by the data in either study. In study 1, the only difference in gaze shifts between the agent (*bear*) and patient (*man*) (or vice versa) reflecting points in time when speakers were uncertain about order of mention and syntactic structure was for the point in time corresponding to the onset of the main subject (*woman*). This pattern did not hold for study 2. As stated before, the fact that speakers did not make use of the structural primes suggests that the manipulation was ineffective. Furthermore it raises the question of whether the current study adequately operationalzed indecision. That is, was it wrong to operationalze indecision as the difference in number of gaze shifts between the agent and patent (or vice versa) in the embedded scene? Nonetheless, the pattern of results for study 1 can be narrowly interpreted as partial support for the *advanced planning* hypothesis. That is, speakers make decisions about syntactic structure prior to or very early during articulation.

Study 2 revealed no statistically significant results and did not support any stated predictions but it did reveal an interesting pattern of results (also present in study 1) that can be interpreted as additional support for previous eye movement studies that found a relationship between eye movement patterns and single word production (e.g., Van Der Meulen, 2001). For example, even in the absence of an effect of the bias manipulation both studies showed an increase in gaze shifts at the clause boundaries (e.g., prior to speech onset and prior to the onset of the embedded clause in study 1 and prior to the onset of the embedded clause in study 2). This suggests, similar to that of Goldman-Eisler (1968) and Ford and Holmes (1978), that syntactic planning *may* occur at the level of the deep structure (meaning) clause, evidenced by an increase in processing (gaze shifts) at the clause boundaries. This claim is also, to a lesser extent, supported by the increase in disfluencies at the clause boundaries (an index of difficulty with word retrieval and/or planning). This interpretation requires a divergence form the initial conceptualization of processing because it ignores uncertainty and only looks at gaze shifts. Nonetheless, it arguably provides partial support for previous eye movement studies.

This interpretation is supported by previous language production studies that have used speech onset latency to investigate eye movement during the production of more complex utterances (e.g., Schriefers & Teruel, 1999). The general finding is that speakers often initiate complex phrases before having planned all their constituents. In terms of eye movements, when speakers produce complex noun phrases, their eyes remain on the referent object until they have fully planned the phrase about the object and are about to initiate the phrase-final word (e.g., Level & Meyer, 2000). This contrasts with earlier eye movement and production studies (e.g., Ferreira, 1991; Meyer 1996) that show an increase in speech onset latencies for syntactically complex utterances (see, Yngve (1960) for a description of *syntactically complex utterances*). These studies suggest that the planning process may be different for the simple and complex utterances which also suggests that it might be necessary to use a combination of syntactically simple and complex utterances to determine if there is a threshold at which the syntactic complexity of an utterance can cause a speaker to shift from *on-line* or *frame-based* production to a set of schemas that best fit the situation or task

Though the results of the current study were inconclusive they point to a well know fact about the study language production; the study of language production is both a theoretical and empirical challenge making it important to continue to search for better methods and constraints to enhance our understanding of the processes and units involved in structuring a spoken utterance.

APPENDIX A

ACTIVE AND PASSIVE QUESTIONS: STUDY 1

Item	Active Questions
1	Is a woman thinking about something chasing someone?
2	Is a boy thinking about something destroying something?
3	Is a boy laughing at something chasing someone?
4	Is a bald man thinking about something chasing someone?
5	Is a punk-rocker thinking about something crushing something?
6	Is a worried woman thinking about something chasing someone?
7	Is a freckled girl thinking about something awaking someone?
8	Is a girl thinking about something striking something?
9	Is a woman thinking about something biting someone?
10	Is a secretary thinking about something kicking someone?
11	Is a politician reading about someone hitting someone?
12	Is a man reading about something hitting something?
13	Is an old man thinking about something attacking someone?
14	Is a reporter talking about something scaring someone?
15	Is a woman looking at something hitting something?
16	Is a mother thinking about something swallowing someone?
17	Is a lady thinking about something running over someone?
18	Is a businesswoman reading about something delivering something?
19	Is a student thinking about something spraying someone?
20	Is a woman thinking about something hitting someone?
21	Is a young girl thinking about something eating someone?
22	Is a girl thinking about something stinging someone?
23	Is a lady thinking about something hitting someone?
24	Is a writer writing about something hitting someone?

Item	Passive Questions
1	Is a woman thinking about someone being chased by something?
2	Is a boy thinking about something being destroyed by something?
3	Is a boy laughing at someone being chased by something?
4	Is a bald man thinking about someone being chased by something?
5	Is a punk-rocker thinking about something being crushed by something?
6	Is a worried woman thinking about someone being chased by something?
7	Is a freckled girl thinking about someone someone being awaken by something?
8	Is a girl thinking about something being struck by something?
9	Is a woman thinking about someone being bit by something?
10	Is a secretary thinking about someone being kicked by something?
11	Is a politician reading about someone being hit by someone?
12	Is a man reading about something being hit by something?
13	Is an old man thinking about someone being attacked by something?
14	Is a reporter talking about someone being scared by something?
15	Is a woman looking at something being hit by something?
16	Is a mother thinking about someone being swallowed by something?
17	Is a lady thinking about someone being run over by something?
18	Is a businesswoman reading about something being delivered by something?
19	Is a student thinking about someone being sprayed by something?
20	Is a woman thinking about someone being hit by something?
21	Is a young girl thinking about someone being eaten by something?
22	Is a girl thinking about someone being stung by something?
23	Is a lady thinking about someone being hit by something?
24	Is a writer writing about someone being hit by something?

APPENDIX B

ACTIVE AND PASSIVE QUESTIONS: STUDY 2

Item	Active Questions
1	Is a basketball player looking at something chasing someone?
2	Is a scientist reading about something scaring someone?
3	Is a schoolteacher talking about something chasing someone?
4	Is an actress complaining about something chasing someone?
5	Is a teacher worrying about something striking someone?
6	Is an electrician looking at something chasing someone?
7	Is a baby laughing at something kicking someone?
8	Is a general smiling at something striking something?
9	Is a baby crying about something biting someone?
10	Is a musician joking about something kicking someone?
11	Is a toddler screaming about something pinching someone?
12	Is a dancer talking about something striking something?
13	Is a nurse watching something attacking someone?
14	Is a football player daydreaming about something scaring someone?
15	Is an artist describing something hitting something?
16	Is a mechanic laughing at something eating someone?
17	Is a performer singing about something running over someone?
18	Is a wrestler joking about something delivering something?
19	Is a waiter shouting about something spraying someone?
20	Is a mechanic screaming about something hitting someone?
21	Is a cowboy crying about something eating someone?
22	Is a cyclist describing something stinging someone?
23	Is a plumber smiling at something hitting someone?
24	Is a student singing about something hitting someone?

Item	Passive Questions
1	Is a basketball player looking at someone being chased by something?
2	Is a scientist reading about someone being scared by something?
3	Is a schoolteacher talking about someone being chased by something?
4	Is an actress complaining about someone being chased by something?
5	Is a teacher worrying about someone being struck by something?
6	Is an electrician looking at someone being chased by something?
7	Is a baby laughing at someone being kicked by something?
8	Is a general smiling at something being struck by something?
9	Is a baby crying about someone being bit by something?
10	Is a musician joking about someone being kicked by something?
11	Is a toddler screaming about someone being pinched by something?
12	Is a dancer talking about something being struck by something?
13	Is a nurse watching someone being attacked by something?
14	Is a football player daydreaming about someone being scared by something?
15	Is an artist describing about something being hit by something?
16	Is a mechanic laughing at someone being eaten by something?
17	Is a performer singing about someone being run over by something?
18	Is a wrestler joking about something being delivered by something?
19	Is a waiter shouting about someone being sprayed by something?
20	Is a mechanic screaming about someone being hit by something?
21	Is a cowboy crying about someone being eaten by something?
22	Is a cyclist describing someone being stung by something?
23	Is a plumber smiling at someone being hit by something?
24	Is a student singing about someone being hit by something?

REFERENCES

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89, 1-47.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 12, 575-586.

Bock, J. K. (1989). Closed class immanence in sentence production. *Cognition, 31,* 163–186.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). San Diego: Academic Press.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47-67.

Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In W.Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 9-50). Norwood, NJ: Ablex.

Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition, 90,* 29-49.

CMU Sphinx Group. (2001). Sphinx-2 [Computer software]. Retrieved November 2, 2001, from http://sourceforge.net/projects/cmusphinx/.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93*,* 283–321.

Dell, G. S. & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287-314. Reprinted (1993) in W. J. M. Levelt (Ed.), Lexical access in speech production (pp. 287-314). Cambridge, MA: Blackwell.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times prepared utterances. *Journal of Memory and Language*, *30*, 210-233.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of Language Production*. Cambridge, MA: MIT Press.

Ferreira, F., & Engelhardt, P. E. (2006). Syntax and production. In M. A. Gernsbacher & M. Traxler (Eds.), *Handbook of psycholinguistics* (pp. 61-91). Oxford, UK: Elsevier Inc.

Ferreira, V. (1996). It is better to give than to donate? Syntactic flexibility in language production. *Journal of memory and language*, 35, 724-755.

Ford, M. (1982). Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 797–827). Cambridge, MA: MIT Press.

Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, 6, 35-53.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), The psychology of learning and motivation (pp. 133-177). New York: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–256). Amsterdam: North-Holland.

Garrett, M. F. (1980). The limits of accommodation: arguments for independent processing levels in sentence production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen and hand* (pp. 263±272). New York: Academic Press.

Garrett, M. F. (1982). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 170-220).

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology.* 10, 96-106.

Goldman-Eisler, F. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech* (pp. 122-132).

Goldman-Eisler, F. (1968). *Psycholinguistics: experiments in spontaneous speech*. London: New York, Academic Press.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82, B1-B14.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274-279.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin and Review, 10,* 603-609.

Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 105–134). New York: Psychology Press.

Jay, T. (2003). *The Psychology of Language.* New York: Prentice-Hall

Kempen, G. & Hoenkamp, E. (1987) An incremental procedural grammar for sentence formulation. *Cognitive Science, 11,* 201-258.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Medin, D. L., Ross, B. H., & Markman, A. B. (2001). Implicit and Explicit Memory. In B. J. Potthoff, P. Howell, & E. Richards (Eds.), *Cognitive Psychology* (pp. 210-220). Orlando, FL: Harcourt College Publishers.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results form picture-word interference experiments. *Journal of Memory and Language*, 35, 477-496.

Meyer, A.S., Sleiderink A.M., & Levelt, W.J.M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66,* B25-B33.

Meyer, A. S., & Van Der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin and Review*, 7, 314-319.

Schriefers, H., Teruel, E., & Meinshausen, R.M. (1998). Producing simple sentences: Results from picture-word interference experiments. *Journal of Memory and Language*, *39*, 609–632.